

Ensaio clínico randomizado: Série Entendendo a Pesquisa Clínica 2

Randomized clinical trials: Series Understanding Clinical Research 2

Marco Aurelio Pinho de Oliveira¹
Luis Guillermo Coca Velarde²
Renato Augusto Moreira de Sá³

Palavras-chave

Ensaio clínico
Experimentação humana
Pesquisa
Técnicas de pesquisa

Keywords

Clinical trial
Human experimentation
Research
Investigative techniques

Resumo

Os ensaios clínicos constituem-se numa poderosa ferramenta para a avaliação de intervenções para a saúde. Na avaliação crítica de um ensaio clínico randomizado, devemos considerar o padrão de recrutamento e seleção, e os critérios de inclusão e exclusão. Trabalhos com critérios muito rígidos de inclusão são mais restritivos na generalização das conclusões. Cabe ao leitor decidir se as diferenças entre a população selecionada no estudo e a população que lhe interessa são importantes ou não.

Abstract

Clinical trials constitute a powerful tool for the evaluation of interventions for health. In a critical analysis of a randomized clinical trial, we must consider the pattern of recruitment and selection, and the inclusion and exclusion criteria. The trials with very strict inclusion criteria are more restrictive in the generalization of the findings. The reader must decide whether the differences between the selected population in the study and the population of interest are important or not.

Estudo realizado na Universidade do Estado do Rio de Janeiro (UERJ) – Rio de Janeiro (RJ), Brasil.

¹Professor Adjunto de Ginecologia da UERJ – Rio de Janeiro (RJ), Brasil.

²Professor Associado do Departamento de Estatística da Universidade Federal Fluminense (UFF) – Niterói (RJ), Brasil.

³Professor Associado de Obstetrícia do Departamento Materno-Infantil da UFF – Niterói (RJ), Brasil.

Endereço para correspondência: Renato Augusto Moreira de Sá – Pós-Graduação em Ciências Médicas, Universidade Federal Fluminense – Avenida Marques de Paraná, 303, 4º andar do prédio anexo – CEP: 24033-900 – Niterói (RJ), Brasil – E-mail: rsa@cpdt.com.br

Conflito de interesses: não há.

O desenho de um estudo pode ser definido como a forma em que se reúnem e comparam pessoas com determinadas características com a finalidade de avaliar a associação entre um fator de risco e uma enfermidade ou entre um tratamento e a evolução desta.

Uma classificação prática é aquela que divide os estudos em dois grandes grupos: experimentais e não experimentais. Os estudos experimentais são sempre analíticos, longitudinais e prospectivos. Dos vários tipos de estudos experimentais, o de uso mais frequente, uma vez que proporciona evidências mais fortes, é o ensaio clínico randomizado (ECR). Diferente dos estudos observacionais em que o pesquisador não interfere na exposição, nesse estudo o pesquisador planeja e intervém ativamente nos fatores que influenciam os indivíduos da amostra. A alocação dos sujeitos de pesquisa pode ser de forma aleatória (randomizada) ou não aleatória. Na sua forma mais simples, os ensaios clínicos são estudos nos quais um grupo de interesse em que se faz uso de uma terapia ou exposição é acompanhado comparando-se com um grupo controle.

Na avaliação de intervenções para a saúde, sejam elas medicamentosas ou não, os ensaios clínicos são de demasiada importância¹. O primeiro ensaio clínico nos moldes que hoje conhecemos foi publicado no final da década de 1940², quando o estatístico Sir Austin Bradford Hill alocou aleatoriamente pacientes com tuberculose pulmonar em dois grupos: os que receberiam estreptomicina e os que não receberiam o medicamento. Dessa forma, ele pode avaliar, de maneira não viesada, a eficácia desse medicamento. Pese a publicação crescente de ensaios clínicos controlados, alguns aspectos do desenho e da análise ainda são mal compreendidos e interpretados de forma equivocada.

Embora os ECRs sejam o padrão-ouro para determinação de efeito de uma terapêutica, por diversos motivos, seu uso na cirurgia não é tão difundido como nas outras especialidades médicas³.

Dentre as principais desvantagens destes estudos, há que se considerar que são laboriosos e custosos, e na cirurgia trazem consigo diversos desafios práticos e metodológicos na sua elaboração e realização. Já dentre as vantagens, a principal, é o fato de que a alocação aleatória neutraliza as prováveis influências dos vieses e fatores de confundimento, além de que podem dar uma descrição completa da experiência emanada pela exposição dos distintos efeitos causais, podendo assim se obter informações de riscos e benefícios. Permitem ainda avaliar o calendário de eventos.

As comparações de intervenções cirúrgicas têm sido realizadas por muitos anos. Várias intervenções cirúrgicas usadas por períodos longos de tempo têm se mostrado ineficazes e sido descartadas após submetidas a esses estudos⁴. A ligadura da artéria mamária interna foi descartada após resultados de dois pequenos ensaios clínicos não terem demonstrado diferença em relação ao placebo.

Mas, surpreendentemente, o número de ensaios clínicos na cirurgia não acompanha o de outras especialidades. Uma revisão encontrou um aumento no número de ensaios clínicos randomizados de 1990 a 2000⁵. No entanto, apenas 3,4% do total de artigos em importantes revistas cirúrgicas foram ensaios clínicos⁶. O contraste com outras áreas pode ser visto por comparações entre auditorias de bases de dados para a prática clínica em medicina interna e diversas especialidades cirúrgicas. Ao considerar que metade das intervenções em medicina interna foi julgada com base em evidências de ensaios clínicos, duas auditorias cirúrgicas relataram um quarto ou menos quando avaliadas evidências cirúrgicas^{6,7}.

O principal problema relacionado aos ensaios clínicos na cirurgia e que dificulta sua confecção e realização é a dificuldade no mascaramento. O mascaramento é o processo de retenção da informação sobre as intervenções atribuídas a cada grupo e é um elemento-chave na concepção de ECRs. Quando feito com sucesso, é reconhecido como tendo um papel importante na prevenção da introdução de um viés de informação. O mascaramento, também conhecido como cegamento, pode ser aplicado aos participantes, aos pesquisadores e aos avaliadores dos resultados do estudo. Em ensaios clínicos cirúrgicos, o mascaramento implica um desafio maior do que para terapias medicamentosas. É impossível impedir que o cirurgião saiba qual técnica ele estará usando. Logo, é impossível o mascaramento do médico. Os avaliadores podem, e devem, preferencialmente, ser mascarados quanto aos resultados. No entanto, o mascaramento não ocorre geralmente em intervenções cirúrgicas para nenhum dos participantes do estudo. Isso ocorre em razão do hábito de o paciente retornar ao médico que o operou para o seguimento. O mascaramento dos pacientes pode ser conseguido quando se usa a mesma via para cirurgias diferentes. Isso é mais difícil de se obter quando as vias são diferentes, por exemplo, laparotômica comparada com laparoscópica, e virtualmente impossível quando se compara cirurgia com medicamentos. Apesar de cirurgia como placebo ser teoricamente possível, motivos éticos impedem sua concepção na prática embora haja relatos na literatura⁸.

Ensaio clínicos bem conduzidos são o padrão-ouro entre os estudos para guiarem a nossa prática clínica diária. A elaboração de um ECR visa a um estudo que seja suficientemente grande para que um efeito clínico importante seja estatisticamente significativo, mas não tão grande para que desfechos pouco importantes sejam significativos. Além disso, quando excessivamente grandes, os ECRs podem ser proibitivamente caros⁹. O tamanho da amostra para qualquer estudo é calculado com base na magnitude das probabilidades dos erros tipo I e II que o investigador está disposto a tolerar. Um ponto forte dos ECRs é que a alocação aleatória evita o fator de confundimento por

dispersar igualmente esses fatores entre os grupos de forma paritária. No entanto, existem importantes inconvenientes de um único ECR de grande porte. Primeiro, alguns ECRs são tão rigorosos na sua metodologia que os resultados podem somente servir para essa população estudada (problemas de validade externa ou de generalização). Além disso, os pacientes que participam de ensaios clínicos podem ter uma linha de base de risco para o desfecho de interesse muito maior do que a média dos pacientes da população ou podem ser oriundos de populações muito diferentes. Por exemplo, alguns dos estudos que estudaram a ingestão de cálcio para a prevenção da pré-eclâmpsia (embora com metodologia rigorosa) foram realizados em países com dietas muito diferentes entre si, o que impede conclusões para populações com outras dietas. Isso pode impedir uma generalização dos resultados. Além disso, alguns ensaios clínicos podem não possuir um número suficiente de indivíduos em subgrupos importantes. Uma preocupação final com o ECR é a questão do viés de publicação. O viés de publicação ocorre quando os resultados dos estudos que relatam resultados positivos (com significância estatística) são mais prováveis de serem publicados do que estudos com resultados negativos. Os estudos que relatam resultados significativos podem resultar em um maior número de publicações e também poderão ser publicados em jornais de maior impacto. Isso pode ter impacto na confecção de metanálises que podem não englobar ensaios com resultados desfavoráveis.

Apesar das vantagens inerentes aos ECRs, esse tipo de estudo deve ser reservado para algumas situações, pois costumam ser mais caros e trabalhosos¹⁰.

1. Ensaios clínicos randomizados devem ser considerados quando:
 - há incerteza sobre o efeito de uma exposição ou tratamento;
 - a exposição pode ser modificada no estudo.
2. Potenciais limitações de ensaios clínicos randomizados incluem:
 - limitada generalização da população em estudo;
 - limitada generalização do ambiente em estudo;
 - Ensaios clínicos randomizados respondem a uma questão de estudo específica.
3. As medidas de magnitude de efeito usuais em ensaios clínicos randomizados são o risco relativo e a diferença de risco.

Em ECRs, um tratamento ou procedimento pode ser comparado a nenhum tratamento, a uma terapêutica semelhante (mesma classe de drogas nos *head-to-head trials*), com um placebo ou a uma norma preexistente de acompanhamento¹⁰. ECRs são realizados sob condições controladas para garantir que a hipótese do estudo é testada de uma forma reproduzível. Protocolos de ensaios clínicos randomizados muitas vezes

não podem ser repetidos na prática clínica. Além disso, caso a amostra do estudo seja muito selecionada, pode não representar a população geral — nesse caso, a generalização pode ficar comprometida. O ambiente de estudo especializado de um estudo randomizado pode resultar em conclusões válidas internamente, mas os resultados não podem ser generalizados para pacientes com a mesma condição fora do estudo¹⁰.

Ensaios clínicos randomizados são projetados para responder definitivamente a uma questão específica, concentrando-se estritamente em uma hipótese de pesquisa, isolando o efeito de uma ou de um pequeno número de terapias. Ensaios clínicos randomizados não são projetados para avaliar os mecanismos pelos quais uma terapia pode produzir benefícios ou prejuízos. Costumam ter uma aplicabilidade clínica limitada. Conclusões retiradas de ECRs são limitadas a situações em que a exposição de interesse pode ser modificada em um ambiente de estudo, por exemplo, uso de medicamentos ou mudanças de estilo de vida. Há inúmeras exposições que são altamente relevantes para estudo, mas não podem ser facilmente modificadas, como genes, marcadores séricos, condições socioeconômicas, etc. Ensaios clínicos randomizados são limitados a situações clínicas específicas nas quais a exposição de interesse pode ser facilmente modificada. Também não podem ser usados em situações nas quais a exposição seja reconhecidamente deletéria, por não ser ético. Por exemplo, não podemos fazer com que alguém fume para compararmos com um grupo que não fume por já sabermos dos malefícios do fumo.

A pesquisa de drogas pela indústria envolve ECRs nas suas diferentes fases. Elas são realizadas basicamente em quatro fases¹⁰:

- Estudos fase I – O primeiro passo no desenvolvimento de uma nova droga é entender se o medicamento é bem tolerado em um pequeno número de pessoas. Embora não seja um ensaio clínico, esses tipos de estudos são referidos como estudos de fase I. Os participantes de estudos de fase I são adultos saudáveis ou pessoas com a doença específica que a droga se destina a tratar. Ocasionalmente, estudos de fase I não podem ser realizados em adultos saudáveis porque a droga tem inaceitáveis efeitos adversos, como agentes quimioterápicos. Estudos de fase I procuram determinar até que dose uma droga pode ser administrada antes de ocorrer toxicidade inaceitável. Esses estudos são iniciados com baixas doses em número limitado de pessoas e, em seguida, aumenta-se a dose gradualmente.
- Estudos de fase II – São projetados para avaliar se um medicamento possui atividade biológica e para determinar sua segurança e tolerabilidade.

- Estudos de fase III/IV – Os estudos de fase III são ensaios clínicos randomizados delineados para avaliar a eficácia e a segurança de uma intervenção. Os resultados dos estudos de fase III são desfechos clínicos, como morte ou sobrevida livre de tumor. As avaliações de segurança ocorrem durante um período mais longo quando comparadas com estudos de fase II. Estudos de fase IV ocorrem após aprovação e avaliam os resultados associados a uma droga ou intervenção na prática clínica com o uso na população geral.

Na avaliação crítica de um ensaio clínico randomizado, devemos fazer algumas perguntas importantes, usando algumas informações já assinaladas anteriormente¹¹.

A primeira pergunta que se deve fazer é “Os pacientes selecionados neste estudo são semelhantes aos pacientes que eu trato?”. O trabalho deve dizer claramente quais foram os critérios de inclusão e exclusão (idade, sexo, doenças anteriores, etc.). Trabalhos com critérios muito rígidos de inclusão são mais restritivos na generalização das conclusões (porém são mais específicos). Deve-se mencionar ainda como os pacientes foram selecionados da população geral (“Só entraram os pacientes que procuraram o hospital terciário?”, “São apenas voluntários remunerados?” ou “São os pacientes de uma comunidade carente próxima ao hospital?”). É importante lembrar que só podemos generalizar para as populações semelhantes a do estudo em questão. Cabe ao leitor decidir se as diferenças entre a população selecionada no estudo e a população que lhe interessa são importantes ou não.

Um dos aspectos fundamentais é o evento final de interesse. Muitos trabalhos são até metodologicamente bem montados; porém, estudam variáveis que não são muito úteis clinicamente ou que não representam aquilo que deveriam. Por exemplo, no tratamento da endometriose com uma medicação antiestrogênica, o pesquisador verifica somente uma diminuição estatisticamente significativa do score da AFS (quantifica a extensão da doença) e conclui que o tratamento é eficaz. Porém, o que a endometriose causa é basicamente infertilidade e dor pélvica, e o estudo em questão não avaliou o que realmente interessa (evento final de interesse), ou seja, se a paciente melhorou da dor ou se conseguiu engravidar (não interessa à paciente melhorar somente o score da AFS). Outro exemplo seria uma pesquisa para verificar a eficácia de um novo quimioterápico, tendo como parâmetro apenas a diminuição da massa tumoral (porém, a diminuição do tumor pode não se correlacionar com a sobrevida, variável de interesse para a paciente).

Outro aspecto importante é saber como foi feita a distribuição dos indivíduos entre o grupo tratamento e o grupo controle (ou placebo). Trabalhos que não incluem grupo

controle ou placebo perdem muita credibilidade (correm o risco de mostrar que um tratamento é eficaz quando na verdade não é), pois o placebo pode produzir efeitos (por exemplo, melhora da cefaleia) superiores a 50% (pode inclusive ser melhor que o medicamento estudado!), ou seja, a melhora do paciente pode acontecer não pelo tratamento em questão, mas sim por outros motivos (história natural da doença, aspectos psicológicos, etc.). Algumas vezes, a utilização de um placebo (não tratamento) é inadequada, como no tratamento de uma doença inflamatória pélvica aguda. Nesses casos, o pesquisador pode comparar um novo medicamento usando como grupo controle o tratamento considerado padrão.

O próximo passo a ser verificado é saber como foi feita a randomização entre os grupos tratamento e controle. A randomização implica a distribuição aleatória dos pacientes, de modo que cada caso possa ter a mesma chance de ser alocado nos diferentes grupos. Estudos controlados (i.e., com grupo controle) não randomizados perdem na qualidade da evidência científica, pois, na maioria das vezes, os grupos não são homogêneos (grupos viciados). Por exemplo, em um estudo comparativo entre a cirurgia de Burch laparoscópica *versus* laparotômica, pode acontecer de serem selecionadas para a laparoscopia pacientes mais magras e com queixas mais leves de incontinência urinária de esforço. Desse modo, não se pode comparar esses dois diferentes grupos de pacientes (nesse caso, desfavorece a cirurgia por laparotomia). A randomização acaba ajustando esses fatores conhecidos (idade, peso, grau da doença, etc.), assim como fatores desconhecidos (por exemplo, algum fator genético desconhecido e que possa alterar os resultados). A randomização adequada deve ser feita com números aleatórios gerados por computador (também existem tabelas apropriadas) que são colocados em envelopes lacrados numerados que são abertos imediatamente antes da aplicação do tratamento. Métodos não adequados de randomização (quasi-randomizados) incluem, por exemplo, fazer cirurgia por laparoscopia às segundas e quartas-feiras e por laparotomia às terças e quintas, ou então fazer uma cirurgia por laparoscopia e outra por laparotomia (alternando as vias). Esses métodos propiciam falhas na distribuição aleatória, mesmo que inconscientemente. Por exemplo, o cirurgião (ou quem encaminhou a paciente para a cirurgia e que conhece o esquema do estudo) pode achar a paciente muito obesa para a laparoscopia e subliminarmente (ou propositalmente?) adiar a cirurgia por um dia, coincidindo com a vez da laparotomia. A randomização adequada evita esse tipo de erro na maioria das vezes. Entretanto, a randomização bem feita não garante a homogeneidade dos grupos (principalmente com amostras

pequenas). É de bom tom verificar antes da análise dos dados se houve realmente uma distribuição semelhante dos pacientes (particularmente, das características que influenciam diretamente no resultado, como: estágio da doença, idade, peso, etc.). Grupos não balanceados podem enviesar os resultados.

Deve-se também estar atento quanto a se a mensuração do evento final de interesse foi realizada sob o mascaramento duplo (duplo-cego). Isso implica que nem o paciente nem o profissional que está fazendo a avaliação (por exemplo, verificar o grau de dor no pós-operatório em pacientes que usaram ou não marcaína na ferida operatória) sabe qual tratamento foi aplicado. O médico pode, inconscientemente, não valorizar as queixas das pacientes que pertencem ao grupo do novo tratamento, falseando os resultados. O próprio paciente pode ser simpático a um determinado tipo de tratamento e não informar ao médico todas as queixas que tem. O método duplo-cego tenta eliminar esse potencial tipo de erro. Nem sempre existe a possibilidade de se conduzir um estudo duplo-cego (comparação entre tratamento clínico *versus* cirúrgico), porém não se deve deixar de utilizá-lo quando possível.

Outro aspecto importante é o tempo de seguimento (*follow-up*). Em primeiro lugar, é preciso certificar-se de que o tempo de seguimento foi o suficiente para avaliar o evento final de interesse. Por exemplo, na transecção laparoscópica dos ligamentos uterossacros, com o objetivo de aliviar a dismenorria, o tempo de seguimento deve ser superior a um ano, pois se sabe do número elevado de recidivas após esse período. Outro aspecto relevante é verificar se o percentual de pacientes que abandonaram o seguimento foi semelhante entre os diversos grupos. Taxa de seguimento inferior a 70% praticamente inviabiliza a análise dos dados, pois o abandono pode ter acontecido por efeitos colaterais, ou até pelo óbito decorrentes do tratamento.

Os erros mencionados anteriormente são sistemáticos, diferente dos erros aleatórios, que podem acontecer quando trabalhamos com qualquer amostra. A análise estatística adequada visa identificar se as diferenças obtidas nos resultados foram por conta dos erros aleatórios ou se realmente existe diferença entre os tratamentos aplicados.

Leituras suplementares

1. Medical Research Council. Streptomycin treatment of pulmonary tuberculosis. A Medical Research Council Investigation. *BMJ*. 1948;2:769-82.
2. Coutinho ESF, Cunha GM. Conceitos básicos de epidemiologia e estatística para a leitura de ensaios clínicos controlados. *Rev Bras Psiquiatr*. 2005;27(2):146-51.
3. Cook J. The challenges faced in the design, conduct and analysis of surgical RCTs. *Trials*. 2009;10:9.
4. Ko CY, Sack J, Chang JT, Fink A. Reporting randomised, controlled trials: where quality of reporting may be improved. *Dis Colon Rectum*. 2002;45:443-7.
5. Wente MN, Seiler CM, Uhl W, Büchler MW: Perspectives of evidence-based surgery. *Dig Surg*. 2003;20:263-9.
6. Ellis J, Mulligan I, Rowe J, Sackett DL: Inpatient general medicine is evidence based. *Lancet*. 1995;364:407-10.
7. Howes N, Chagla L, Thorpe M, McCulloch P. Surgical practice is evidencebased. *Br J Surg*. 1997;84:1220-3.
8. Kenny SE, Shankar KR, Rintula R, Lamont GL, Lloyd DA. Evidence-based surgery: interventions in a regional paediatric surgical unit. *Arch Dis Child*. 1997;76:50-3.
9. Scifres CM, Iams JD, Klebanoff M, Macones GA. Metaanalysisvs large clinical trials: which should guide our management? *Am J Obstet Gynecol*. 2009;200:484.e1-e5.
10. Kestenbaum B. *Epidemiology and Biostatistics: An Introduction to Clinical Research*. Seattle, USA: Springer Science Business Media; 2009.
11. Oliveira MAP, Camara RCM. Entendendo ensaios clínicos randomizados. *Bras J Video-Sur*. 2010;3(4):176-80.